

Reconstructing the galaxy redshift distribution from angular cross power spectra

Lei Sun^{1,2*}, Hu Zhan² and Charling Tao¹

¹*Physics Department and Tsinghua Center for Astrophysics (THCA), Tsinghua University, Beijing, 100084, China*

²*Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The control of photometric redshift (photo- z) errors is a crucial and challenging task for precision weak lensing cosmology. The spacial cross-correlations (equivalently, the angular cross power spectra) of galaxies between tomographic photo- z bins are sensitive to the true redshift distribution $n_i(z)$ of each bin and hence can help calibrate the photo- z error distribution for weak lensing surveys. Using Fisher matrix analysis, we investigate the contributions of various components of the angular power spectra to the constraints of $n_i(z)$ parameters and demonstrate the importance of the cross power spectra therein, especially when catastrophic photo- z errors are present. We further study the feasibility of reconstructing $n_i(z)$ from galaxy angular power spectra using Markov Chain Monte Carlo estimation. Considering an LSST-like survey with 10 photo- z bins, we find that the underlying redshift distribution can be determined with a fractional precision ($\sigma(\theta)/\theta$ for parameter θ) of roughly 1% and 10% for the mean redshift and width of $n_i(z)$, respectively.

Key words: cosmology: observations – cosmology: theory – large scale structure of Universe

1 INTRODUCTION

Weak gravitational lensing is considered a powerful cosmological probe, and much work has been done in the past few decades to advance its measurement technique, its application in cosmology, and our understanding of its systematics (e.g., Bacon et al. 2001; Heymans et al. 2006; Miller et al. 2007; Bridle et al. 2009; Hoekstra et al. 2002; Jarvis et al. 2003; Rhodes et al. 2004; Heymans et al. 2004; Hoekstra et al. 2006; Semboloni et al. 2006; Massey et al. 2007; Benjamin et al. 2007; Kilbinger et al. 2013; Jee et al. 2013; Liu et al. 2015; Fu et al. 2014; Köhlinger et al. 2015; Fan 2007; Yu et al. 2015; Joachimi et al. 2015). With the success of precursor projects (e.g., CTIO, COSMOS, DLS, CFHTLenS), several more ambitious weak lensing surveys targeting at hundred to thousand times larger areas, such as DES¹, LSST², and EUCLID³, are ongoing or under construction. Dark energy is a key science driver of these surveys. In the meantime, various systematic effects (see reviews by Refregier 2003, Bartelmann 2010 and Kilbinger 2015) can become dominant contaminations as statistical er-

rors are expected to reduce to insignificant levels for future surveys. The necessity of obtaining vast amount of galaxy redshifts in weak lensing observations makes the multi-color photometric determination of redshifts the only practical means available. The photometric redshift (photo- z) errors arising therefrom are among the most critical systematics. They are often characterized by the bias and scatter of the photo- z z_p with respect to the spectroscopic redshift (spectro- z) z_s , as well as the so-called catastrophic errors, which can be loosely defined as the outliers with $|z_p - z_s| > 1$. Photo- z errors contaminate the cosmological information in weak lensing and severely degrade the constraints on dark energy (e.g., Huterer et al. 2006; Ma et al. 2006; Zhan 2006; Amara & Réfrégier 2007; Abdalla et al. 2008; Sun et al. 2009; Hearin et al. 2010, 2012). In order to fully exploit the potential of future weak lensing surveys like the LSST, uncertainties on both the bias and the scatter of z_p have to be controlled to below 0.003 (e.g., Huterer et al. 2006; Ma et al. 2006), which is a stringent requirement.

Recently, a technique is developed to calibrate the true redshift distribution of the photo- z sample via cross-correlations with a spectro- z sample in the same volume (Lima et al. 2008; Cunha et al. 2009; Matthews & Newman 2010; McQuinn & White 2013; Schmidt et al. 2013; Ménard et al. 2013; Rahman et al. 2015; Eriksen & Gaztañaga 2015). The spectro- z sample

* E-mail: sunl@nao.cas.cn

¹ <http://www.darkenergysurvey.org>

² <http://www.lsst.org>

³ <http://sci.esa.int/euclid/>

serves as an accurate redshift reference. As one slices through redshift space, the cross-correlation between the spectro- z sub-sample around z and the photo- z sample increases with the overlap between the two and thus maps the true redshift distribution of the photo- z sample. It is still challenging, however, to obtain a good sample of galaxy spectra covering a sufficiently large area to moderately high redshift ($z \sim 4$) for deep surveys like LSST. Moreover, such a calibration is normally valid only in the volume where the photo- z and spectro- z samples overlap. To extend over the whole photo- z survey area, one has to assume that the difference between the photo- z sample inside and that outside the calibration area is negligible.

Separately, it is noted that the galaxy angular auto power spectra can provide useful consistency test on the photo- z error distribution (Zhan & Knox 2006). Zhan (2006) further shows with the simple parametrization⁴ of a photo- z bias and an rms error per redshift interval (not per bin) $\Delta z = 0.1$ that the photo- z error distribution can be calibrated to satisfactory precision by the auto and cross power spectra of the very same galaxies in the weak lensing survey. As a result, the joint analysis of galaxy clustering (the galaxy angular power spectra) and weak lensing is much less prone to uncertainties in the photo- z bias and rms parameters. Schneider et al. (2006) also carries out a study, using a Fisher Matrix analysis, on how well the redshift distribution of galaxies in 6 bins can be determined by angular power spectra of the 6 bins alone. The results show that if the galaxies in each bin are allowed to be wrongly assigned to all other bins with no restriction, the degrees of freedom in the photo- z error distribution would overwhelm the constraining power of the galaxy angular power spectra. Therefore, one should model the photo- z error distribution with as few parameters as possible at an acceptable accuracy. Zhang et al. (2010) combines the angular power spectra and weak lensing analysis to perform the self-calibration, based on Fisher matrix analysis as well. They find that lensing-galaxy correlations also help to improve the photo- z self-calibration by breaking the degeneracy between up scatters and down scatters of photo- z samples. Benjamin et al. (2010) proposes to parameterize the redshift distributions with the total sample number N_i and the contamination fraction from the i th to the j th bin as f_{ij} . With this parametrization, the authors investigate the feasibility of reconstructing N_i and f_{ij} of each bin in a 4-bin case both from simulated and real data.

In this paper, we focus on the reconstruction of the true redshift distribution of each photo- z bin in tomographic weak lensing analysis, using angular cross power spectra. We first build a simple double-Gaussian model to characterize the true redshift distribution of galaxies within tomographic photo- z bins, which is motivated by the bimodal distribution of photo- z errors in presence of catastrophic errors. This model can also be extended straightforwardly to a multi-Gaussian form in the case of multiple catastrophic fractions. Based on this model, we analyze the contributions of various components of the galaxy angular power spectra, such as the cross power spectra, the baryon acoustic oscillations (BAO) and the broadband shape, to the reconstruction

of the underlying redshift distribution. The performance of this calibration method in practical data analyses is investigated with 10 tomographic photo- z bins, utilizing simulated data and a Markov Chain Monte Carlo (MCMC) estimate (e.g., Gilks et al. 1996). The simple double-Gaussian model suffices the purpose of this work, i.e., demonstrating the feasibility of this self-calibration method. The precise modeling of the galaxy true z -distribution in a given photo- z bin, however, will be crucial for the control of photo- z errors in future weak lensing surveys. Thus we plan to perform directly a principle component analysis (PCA) to characterize the probability distribution function $P(z_s|z_p)$.

The rest of this paper is arranged as follows. In §2, the galaxy angular power spectra and the redshift distribution model used in our study is introduced. In §3, we analyze the contributions of various components on galaxy angular power spectra to the reconstruction of the true redshift distribution. Using MCMC estimate, we study the feasibility of this reconstruction method in practical analysis in §4. A summary is presented in §5.

2 METHOD

We first describe how the galaxy angular power spectra is calculated theoretically in our study. We then present the (double-)Gaussian model that is used to characterize the galaxy true redshift distribution in each tomographic photo- z bin.

2.1 Galaxy Angular Power Spectra

The galaxy angular power spectrum, under the Limber approximation (Limber 1954), can be written as (Hu & Jain 2004)

$$P_{ij}(\ell) = \frac{2\pi^2}{\ell^3} \int dz H(z) D_A(z) W_i(z) W_j(z) \Delta_m^2(k; z), \quad (1)$$

where the subscripts i, j label the tomographic bins. $H(z)$ and $D_A(z)$ are the Hubble parameter and comoving angular diameter distance, respectively. $\Delta_m^2(k; z)$ denotes the dimensionless matter power spectrum, where $k = \ell/D_A(z)$ is the wavenumber that projects onto multiple ℓ at redshift z . $W_i(z)$ is the weighting function, $W_i(z) = b(z)n_i(z)$, with $b(z)$ the linear galaxy bias and $n_i(z)$ the normalized (to 1) galaxy redshift distribution of the i th tomographic bin (hereafter denoted as bin- i). We discuss the modeling of $n_i(z)$ specifically in Section 2.2.

The nonlinear matter power spectrum can be obtained by:

$$P_{nl} = (P_{lin} - P_{nw}) \exp\left(-\frac{k^2}{2k_*^2}\right) + P_{nl,nw}, \quad (2)$$

where the first half processes damping of the BAO signal because of nonlinear evolution, with k_* being the characteristic damping scale (Eisenstein et al. 2007). P_{lin} , P_{nw} and $P_{nl,nw}$ are the linear matter power spectrum, its no-wiggle approximation (Eisenstein & Hu 1998) and the nonlinear corrected no-wiggle power spectrum, respectively.

Throughout this study, we consider a fiducial survey resembling the LSST design, with 20,000 square degrees and a redshift range of $0 \leq z \leq 4$. Following Zhan et al. (2006),

⁴ It is still considerably more flexible than a mean redshift and an rms error per photo- z bin.

we model the overall redshift evolution of the galaxy bias as $b(z) = 1 + 0.84z$. Within each bin the galaxy bias is a constant and its fiducial value is assumed at the mean redshift z_m of the individual bin. The galaxy bias parameters then float freely (not bound to the floating z_m), and will finally be marginalized over in the following specific calculations, with a relative 20% prior.

We discard information from very large scales, $\ell < 40$, uniformly for all tomographic bins to avoid potential large-scale effects that are not included in this study (e.g., dark energy clustering). A high multipole cut for each galaxy bin is also incorporated to reduce contaminations of the small-scale non-linearity and baryonic effects. According to Zhan (2006), the maximum multipole for a galaxy bin centred at z_m follows approximately $\ell_{\max} = 340z_m + 346z_m^2$ at $0.15 \leq z_m \leq 0.9$, $446 - 658z_m + 908z_m^2$ at $0.9 < z_m \leq 2.1$, and 3000 at $z_m > 2.1$. Roughly speaking, we make use of the information of galaxy angular spectra on scales $40 \leq \ell \leq 3000$.

The redshift distortion (RSD) effect is not modeled here. It is largely erased in the angular power spectra due to projection along the line of sight, provided that the photo- z bins are wide enough in redshift. In addition, the cuts at low(high)- ℓ ends mentioned above should also help in reducing the impact of the linear RSD on large scales and the 'Finger of God' effect on small scales, respectively.

We consider a flat Λ CDM model with cosmological parameters $w = -1$, $\Omega_m = 0.27$, $\Omega_b = 0.0446$, $n_s = 0.96$, $h = 0.72$ and $\sigma_8 = 0.78$. Throughout, the cosmological parameters are fixed at their fiducial values. The ultimate goal of the self-calibration scenario is to combine the weak lensing and galaxy clustering analysis together to constrain both cosmological and redshift distribution parameters simultaneously. With angular power spectrum alone, here we shall focus on its performance in constraining redshift distribution parameters.

2.2 Modeling the galaxy redshift distribution

For future large sky weak lensing surveys, precise reconstruction of the source galaxy redshift distribution from the defective photo- z measurements is a critical and challenging task. Lots of efforts are made to better model the true redshift distribution $n_i(z)$, given a tomographic photo- z bin i . In previous studies of photo- z calibration using cross correlation techniques, $n_i(z)$ is usually parametrized by a total galaxy number N_i with catastrophic contamination fractions $f_{ij, j \neq i}$, i.e., fraction of galaxies mis-assigned to bin- i from bin- j (e.g., Schneider et al. 2006; Benjamin et al. 2010; Zhang et al. 2010). In this case, $n_i(z)$ is shaped as a histogram. With finer binning within bin- i (correspondingly more parameters), a smooth curve of $n_i(z)$ can also be expected.

This work emphasizes effectiveness and feasibility of the self-calibration method, and we simply adopt a double-Gaussian to model $n_i(z)$, with one Gaussian for the main sample and the other for a sub-sample due to catastrophic failures. This model is inspired by the fact that the confusion of the 4000 break with the Lyman break in galaxy SEDs, the dominant source of catastrophic failures, often leads to a true redshift distribution of bi-model pattern in the relevant tomographic photo- z bins. If there exist various catastrophic failure fractions within one bin, this model

can also be extended straightforwardly to a multi-Gaussian form.

Specifically, the double-Gaussian distribution of bin- i can be written as:

$$n_i(z) = (1 - f^c) \exp \frac{(z - z_{mi})^2}{\sigma_i^2} + f^c \exp \frac{(z - z_{mi}^c)^2}{\sigma_i^{c2}}, \quad (3)$$

where f^c is the fraction of catastrophic failures with respect to the total sample of bin- i . z_{mi} and σ_i denote the mean redshift and the width of the distribution of the main sample, whereas z_{mi}^c and σ_i^c are analogues for the sub-sample of catastrophic failure.

3 CONTRIBUTIONS OF VARIOUS COMPONENTS OF THE ANGULAR POWER SPECTRA

Using Fisher matrix analysis, we clarify the contributions of various components of the galaxy angular power spectra, such as cross(auto) power spectra and BAO, to the constraints of galaxy redshift distribution parameters.

3.1 Parameters

As mentioned in Section 2.1, our fiducial survey is a LSST-like half sky survey with a redshift range $0 \leq z \leq 4$. We consider dividing the whole redshift range into 5 tomographic bins for clarity here, whereas thinner binning can be performed in practical analyses.

As a first step, we examine the case free of catastrophic contaminations. The specific values of redshift distribution parameters z_{mi}, σ_i of each bin are listed in Table 1, which are determined from the overall galaxy redshift distribution in Zhan (2006). The black solid lines in Fig. 1 illustrate the corresponding distribution functions $n_i(z), i = 1, \dots, 5$, each normalized according to the proportion of their surface number density n_g^i to the total surface number density n_g^{tot} . Here, the total surface number density of sampled galaxies is taken to be 40 arcmin^{-2} , and n_g^i of each individual bin (see the 5th column of Table 1.) is set by integrating the radial selection function of LSST (Zhan 2006). Note that the adjacent bins are set overlapped in our model, which naturally mimics the underlying distributions of practical photo- z samples to a certain extent. The cross power spectra are determined by the overlap between bins in true redshift space.

We next include the catastrophic photo- z errors by considering that bin-4 suffers a fraction of $f^c = 10\%$ contamination from bin-1. The underlying distribution of $n_4(z)$ is thus replaced by the double-Gaussian model, shown as the combination of the two red dash lines in Fig. 1. $z_{m4}^c = 0.5$ and $\sigma_4^c = 0.1$ are the mean redshift and width of the catastrophic sub-sample, respectively. The parameters of the main sample of bin-4, (z_{m4}, σ_4), are kept unchanged (only normalization changed to $1 - f^c$).

Table 1. Redshift distribution and survey related parameters. z_m , σ , ℓ_{\max} and n_g are the mean redshift, width, maximum ℓ and galaxy number density of each bin.

bin	z_m	σ	ℓ_{\max}	n_g/arcmin^2
1	0.33	0.14	136	8.92
2	0.74	0.14	425	15.58
3	1.29	0.18	1081	10.69
4	2.01	0.24	2785	3.50
5	2.96	0.32	3000	0.53

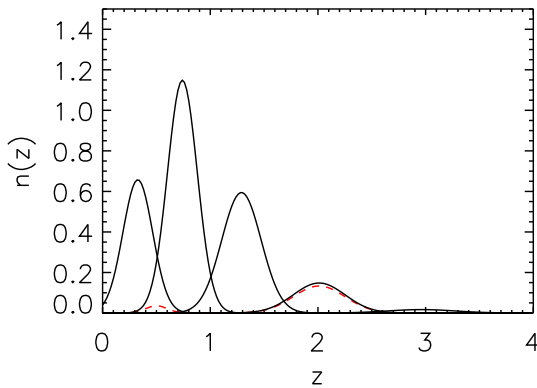


Figure 1. Galaxy redshift distribution curves of the 5 bins. Black solid lines denote those with no catastrophic case. The corresponding z_{mi} , σ_i s are given in Table 1. The 2 red dashed lines present the true distribution of galaxies assigned to bin-4, when there is a 10% contamination of catastrophic failures from bin-1.

3.2 The Fisher Matrix

The Fisher matrix takes the form (Tegmark 1997; for this derived form see also explanations in Zhan 2006)

$$F_{ij} = \sum_{\ell} \left(\frac{\partial C_{ij}(\ell)}{\partial \theta_i} \right)^T \text{Cov}_{ij,mn}^{-1}(\ell, \ell') \frac{\partial C_{mn}(\ell')}{\partial \theta_j}, \quad (4)$$

where the total angular power spectrum $C_{ij}(\ell)$ is given by

$$C_{ij}(\ell) = P_{ij}(\ell) + \delta_{ij} \frac{1}{n_g^i}. \quad (5)$$

$\text{Cov}_{ij,mn}^{-1}(\ell, \ell')$ is the inverse of the covariance matrix. Considering only shot noise and the Gaussian sample variance, it can be written as

$$\begin{aligned} \text{Cov}[C_{ij}(\ell), C_{mn}(\ell')] &= \frac{\delta_{\ell\ell'}}{(2\ell+1)f_{\text{sky}}\Delta\ell} \\ &\times [C_{im}(\ell)C_{jn}(\ell) + C_{in}(\ell)C_{jm}(\ell)]. \end{aligned} \quad (6)$$

Here $\Delta\ell = 1$ is the band width, and $f_{\text{sky}} \approx 0.485$ is the fractional sky coverage of the fiducial survey.

Beyond the cross power spectra, we are also interested in the role of BAO, a minute feature in the power spectrum that is used as a standard ruler to measure angular diameter distances, in constraining $n(z)$. We thus consider 3 cases in which the power spectrum in equation (5) is modified

as follows: 1) the 'auto-P' case with cross power spectra abandoned; 2) the 'no-wiggle' case with BAO wiggles edited out; 3) the 'only-wiggle' case with the broadband shape of the power spectra divided out.

The 'auto-P' case is simply obtained by keeping only the auto power spectra within each z -bin and removing completely the cross correlation information between bins. A comparison with the 'full-P' case, which includes the full information of tomographic angular power spectra, can thus demonstrate clearly the role of cross power spectra. In the 'no-wiggle' case, the Fisher matrix is computed using the nonlinear corrected no-wiggle matter power spectrum ($P_{nl,nw}$, see Section 2.1). In the 'only-wiggle' case, the broadband shape of the power spectrum is fitted with a 3rd order smooth polynomial and then divided out, whose dependence on $n(z)$ parameters is consequently removed from the Fisher matrix. By comparisons of the 'no(only)-wiggle' cases versus the 'full-P' case, the effect of BAO can then be isolated.

3.3 Results

We first derive constraints on the redshift distribution parameters when the catastrophic failure fractions are ignored (the 'no-cata' case). Fig. 2 shows the $1-\sigma$ error contour of $z_{mi}-\sigma_i$ of each bin for each considered situation, with a fractional 20% prior applied on galaxy bias b_i and the other cosmological parameters fixed at fiducial values. The 2-D contour is then compressed to a Figure of Merit (F.o.M)⁵ in Fig. 3 for a quantitative comparison.

By comparing the 'auto-P'(red dotted lines) with the 'full-P'(black solid lines) results from Fig. 2, a significant enlargement of the error region is noticed for all 5 bins when the cross power spectra are discarded. Furthermore, the bin width parameters σ_i rather than the mean redshift z_{mi} gain mostly from the inclusion of the cross power spectra. This implies that the overlapping of adjacent bins caused by wider tails dominates the cross correlation signals. Fig. 3 shows that the omission of the cross power spectra (red bars) can generally result in 45%-75% degradation of the F.o.M of $z_{mi}-\sigma_i$ for different bins. It is confirmed that the cross power spectra play a crucial role in constraining the parameters of redshift distributions.

The BAO feature is also helpful in constraining the redshift distribution parameters, unveiled by the contrast of the 'no-wiggle' with the 'full-P' cases. The blue dashed contours in Fig. 2, which are slightly loosened on both dimensions compared to the solid ones, indicate that the BAO wiggles contribute almost evenly to constraints of z_{mi} and σ_i . The capability of angular (cross) power spectra in self-calibrating the redshift distributions would be ruined if it was a strict power law in ℓ due to degeneracies between the $n_i(z)$ parameters and the normalizations (Zhang et al. 2010). Hence the BAO wiggles help in this regard. Specifically, for bins centered at different z_m and/or with different widths, the BAO wiggles are projected at different ℓ . Removing the BAO wiggles gives rise to a ~ 30 -40% degradation of the F.o.M of $z_{mi}-\sigma_i$ as shown in Fig. 3, though it affects less than removing the cross power spectra does.

⁵ F.o.M = $\sqrt{\det(F)}$, where F is the reduced Fisher matrix for z_{mi} and σ_i .

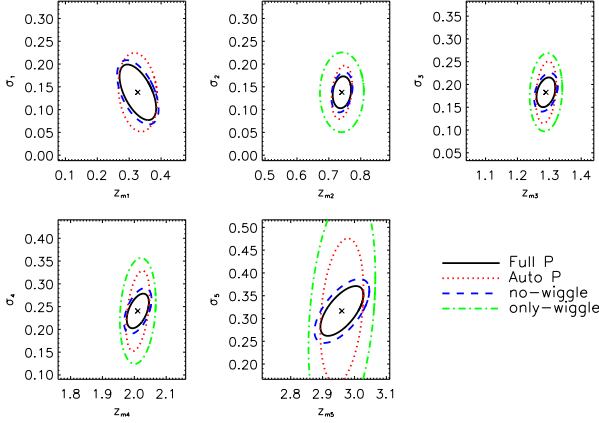


Figure 2. The redshift distribution parameter constraints from various components of the galaxy angular power spectra in the 'no-cata' case. Contours denote the $1 - \sigma$ credible regions.

Discarding the broadband shape of the power spectra, i.e., the 'only-wiggle' case (green dash-dotted lines in Fig. 2 and green bars in Fig. 3), induces a big loss of constraining power. The F.o.M in this case decreases more than 80%. The inconsistency between 'no-wiggle' and 'only-wiggle' results is mainly due to the fact that the division operation in the 'only-wiggle' case induces a further loss of the absolute amplitude of the BAO. Note that due to the high- ℓ cut at $\ell < 136$, there is nearly no BAO signal counted in bin-1, therefore no constraints reported from it for the 'only-wiggle' case.

Next we consider the case with the assumed 10% catastrophic fraction in bin-4 (the 'with-cata' case). The $1 - \sigma$ error region of $z_{mi} - \sigma_i$ of each bin and the corresponding F.o.Ms are shown in Fig. 4 and 5, respectively. First of all, the constraints on $z_{mi} - \sigma_i$ of the main sample of bin-4 and of the other bins are not affected significantly by the considered catastrophic contamination. In the 'full-P' case for instance, the F.o.M of the 'with-cata' case shows a $\sim 15\%$ decrease for bin-4 main sample and $\leq 5\%$ for the remaining bins, compared to that of the 'no-cata' case. Concerning the catastrophic fraction related parameters f^c , z^c and σ^c , the constraints are undoubtedly sensitive to whether the cross power spectra are included or not, since the catastrophic contamination is the main source of the cross correlation between distinct bins. As a result, the 'auto-P' case gives nearly no constraints on these three parameters, with corresponding F.o.Ms $\rightarrow 0$. Besides, as the contamination is originally from the lowest bin-1, the absence of BAO wiggles in the cross correlation between bin-cata and bin-1 nullifies the 'only-wiggle' case in constraining f^c , z^c and σ^c .

4 FEASIBILITY OF RECONSTRUCTING TRUE REDSHIFT DISTRIBUTIONS USING ANGULAR POWER SPECTRA

So far the analysis is based on Fisher matrix, namely, theoretical calculations around the fiducial value. In this section we focus on MCMC estimates of the underlying redshift distribution parameters using simulated angular power spectra

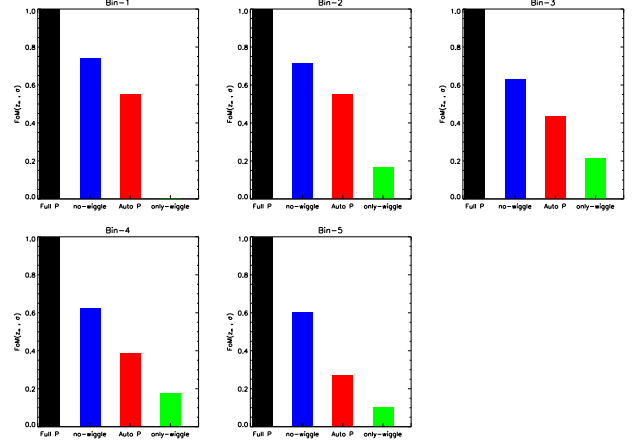


Figure 3. Relative F.o.M of $z_{mi} - \sigma_i$ for each bin in the 'no-cata' case, with the Full-P results normalized to 1.

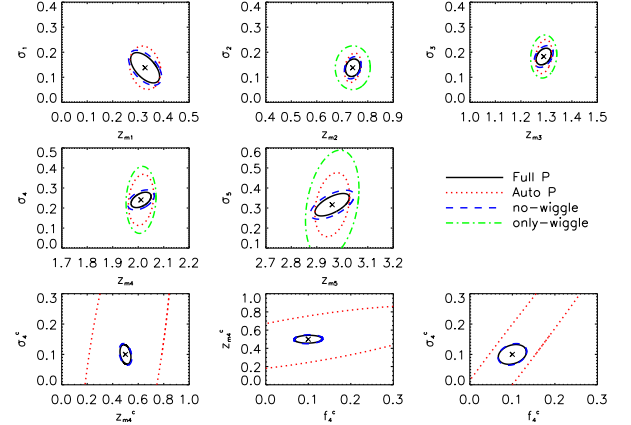


Figure 4. The redshift distribution parameter constraints from various components of the galaxy angular power spectra in the 'with-cata' case. Contours denote the $1 - \sigma$ credible regions.

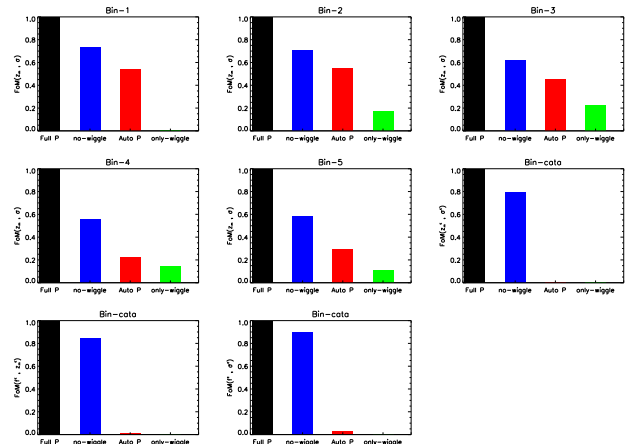


Figure 5. The same as Fig. 3 but for the catastrophic case.

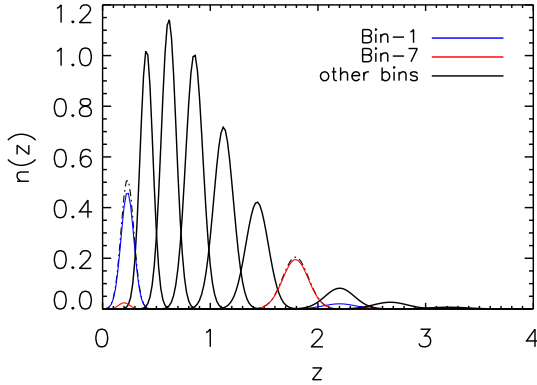


Figure 6. Galaxy distribution curves of the 10 bins in true redshift space. Black solid lines represent bins without catastrophic errors. The corresponding z_{mi}, σ_i s are given in Table 2. The blue(red) lines denote the true distributions of galaxies assigned to bin-1(7), when there is a 10(5)% contamination of catastrophic failures, respectively.

data, aiming to investigate the performance of this method in a practical data analysis.

4.1 Parameters

We consider a more realistic case with 10 tomographic redshift bins, among which bin-1 and 7 are contaminated by catastrophic errors. The two catastrophic error parts are chosen based on the LSST photo-z simulations, which shows two prominent clumps around the corresponding positions in the $z_p - z_s$ plot (Fig. 3.16, [LSST Science Collaboration et al. 2009](#)). While information extracted from the weak lensing analysis tend to be saturated more rapidly with increasing bin number due to the broad kernels, thinner binning usually does help the 2-D galaxy clustering analysis (e.g., angular power spectra) to recover more 3-D information along the line of sight, and consequently a more precise redshift distribution. Certainly, thinner binning means more free parameters, hence a balance has to be found for specific samples.

The specific distributions of each bin in true redshift space are shown in Fig. 6. The combination of two blue solid lines denotes the distribution of bin-1, including its main sample at $z_m \sim 0.2$ and $f^c = 10\%$ catastrophic fraction around redshift 2.2 (denoted as bin-1^c). Similarly, the two red solid lines show the distribution of bin-7, with main sample at $z_m \sim 1.8$ and 5% catastrophic fraction around redshift 0.2 (denoted as bin-7^c). The black solid lines represent distributions of the other bins. The corresponding parameters of each bin, listed in Table 2, are regulated according to the same rules in Section 3.1.

4.2 MCMC estimate and the Likelihood Function

The MCMC method estimates parameters through mapping of the posterior distributions. According to Bayes' Theorem, one can infer $\mathcal{P}(\theta_i | \hat{P}_{ij})$, the posterior probability of

Table 2. Redshift distribution and survey related parameters for the 10-bin case. z_m , σ , ℓ_{max} and n_g are the mean redshift, width, maximum ℓ and galaxy number density of each bin. bin-1^c and 7^c represents the catastrophic sub-samples of bin-1 and 7, respectively, and their fractions within each corresponding bin are given by f^c .

bin	z_m	σ	ℓ_{max}	n_g/arcmin^2	f^c
1	0.23	0.06	90	2.98	10%
1 ^c	2.20	0.15	—	—	—
2	0.41	0.06	190	5.95	—
3	0.61	0.07	330	7.77	—
4	0.85	0.08	540	7.82	—
5	1.12	0.09	860	6.39	—
6	1.43	0.10	1350	4.31	—
7	1.79	0.12	2190	2.40	5%
7 ^c	0.20	0.05	—	—	—
8	2.20	0.14	3000	1.10	—
9	2.67	0.16	3000	0.41	—
10	3.21	0.18	3000	0.12	—

parameters θ_i (e.g., z_{mi} and σ_i), given measured galaxy angular power spectra \hat{P}_{ij} from the likelihood function, once the prior $\mathcal{P}(\theta)$ is specified:

$$\mathcal{P}(\theta | \hat{P}_{ij}) = \mathcal{L}(\hat{P}_{ij} | P_{ij}(\theta))\mathcal{P}(\theta), \quad (7)$$

Concerning the likelihood, the correlation between angular power spectra of different redshift bins in a tomographic case gives rise to a multi-variate Gamma distribution, with a parameter-dependent covariance matrix. Its exact form is computationally unfeasible in practical estimation. A good approximate likelihood is therefore required. Based on discussion in [Sun et al. \(2013\)](#), we choose to use the approximation of multi-variate Gaussian likelihood with its determinant term discarded,

$$\ln \mathcal{L} \propto [P_{ij}(\ell) - \hat{P}_{ij}(\ell)] \text{Cov}_{ij,mn}^{-1}(\ell, \ell') [P_{mn}(\ell') - \hat{P}_{mn}(\ell')]. \quad (8)$$

With cosmological parameters fixed at their fiducial values, it amounts to 38 dimensions in parameter space to explore in the MCMC analysis, i.e., z_{mi}, σ_i and the galaxy bias parameter b_i with $i = 1, \dots, 10$ and $z_{m1(\tau)}^c, \sigma_{1(\tau)}^c, b_{1(\tau)}^c$ and $f_{1(\tau)}^c$ for the catastrophic sub-samples of bin-1(7), respectively. Recall that a fractional 20% prior is applied on b_i . We impose relatively conservative priors on all bins,

$$\begin{aligned} \sigma_P(z_{mi}) &= \sigma_i / \sqrt{N_{\text{spec}} \times (1 - f_i^c)} \\ \sigma_P(\sigma_i) &= \sigma_P(z_{mi}) / \sqrt{2} \\ \sigma_P(z_{mi}^c) &= \sigma_i^c / \sqrt{N_{\text{spec}} \times f_i^c} \\ \sigma_P(\sigma_i^c) &= \sigma_P(z_{mi}^c) / \sqrt{2}, \end{aligned} \quad (9)$$

by taking $N_{\text{spec}} = 9$. This translates to 9 sampled spectra in each bin for calibration, in the context of Gaussian photo-z errors, i.e., regarding parameters z_{mi} and σ_i as the photo-z bias (equivalently) and rms errors. Given the Gaussian mean and scatter, their errors are subject to a $\sqrt{2}$ -times relation (e.g., [Ahn & Fessler 2003](#)) as in the 2nd and 4th lines of equation (9), which we simply take advantage of here.

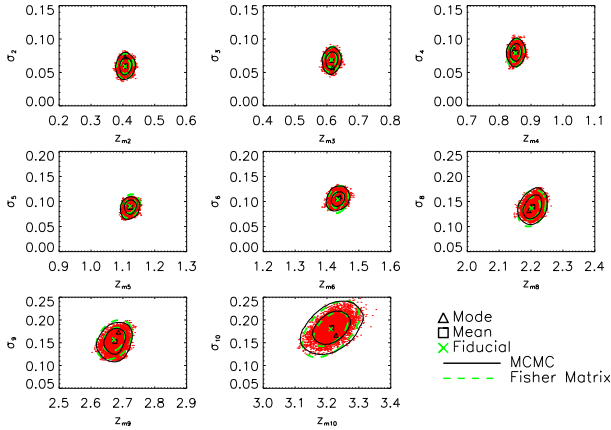


Figure 7. The marginalized constraints on $z_{mi} - \sigma_i$ for bins free of catastrophic errors (bin2-6 and 8-10). The black solid contours indicate the 1,2- σ credible regions estimated from the MCMC samples (red dots, thinned to leave ~ 3000 points for display), with the mode and mean values marked by triangles and squares, respectively. The green dashed contours denote predictions by a Fisher matrix analysis centered at the fiducial values (green crosses).

Provided the survey conditions and fiducial cosmological models, the angular power spectra are calculated theoretically using equation (1) and are subsequently regarded as the ‘observed’ data, $\hat{P}_{ij}(\ell)$. The covariance matrix, including both shot noise and cosmic variance, is modeled following equation (6). We next use the CosmoMC package⁶ (Lewis & Bridle 2002), with slight modifications to perform the MCMC analysis. We run four chains with about 10^5 samples per chain after burn-in and merge them into one sample. The Fisher matrix results are also incorporated as a reference.

4.3 Results

The marginalized constraints on redshift distribution parameters are presented in Fig. 7 & 8 for the eight ‘no-cata’ bins and two ‘with-cata’ bins, respectively. The black solid contours indicate the 1- σ and 2- σ regions estimated from the MCMC samples (red dots, thinned to leave ~ 3000 points for display), with the mode and mean values marked by triangles and squares, respectively. The green dashed contours denote predictions by a Fisher matrix analysis centered at the fiducial values (green crosses).

From Fig. 7, it is evident that for those without-cata bins both the best-fit values (mean or mode) of parameters estimated from MCMC and the errors are well consistent with the input fiducial values and the error level predicted by Fisher matrix analysis, respectively. Roughly speaking, a relative precision ($\sigma(\theta_i)/\theta_i$) of $\sim 1\% - 2\%$ on z_{mi} and $\sim 10\%$ on σ_i can be achieved. The absolute errors in z_{mi} vary from ~ 0.01 to ~ 0.03 with increasing z_{mi} , and $\sim 0.007 - 0.02$ in σ_i .

For the ‘with-cata’ bins shown in Fig. 8, parameters of the main sample distribution of bin-1 get less constrained

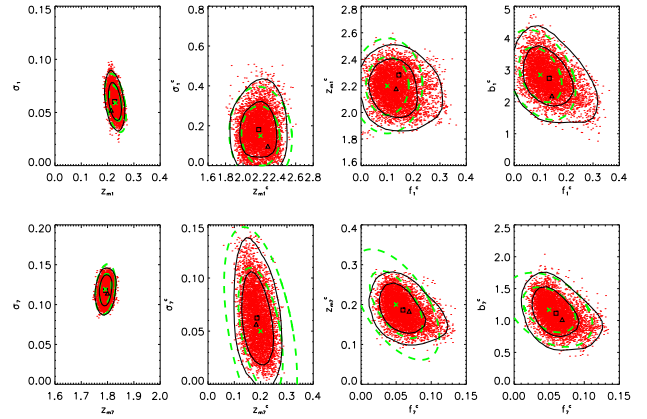


Figure 8. The same as Fig. 7 but for parameters of bin-1 (upper panels) and bin-7 (lower panels).

with fractional errors 6% on z_{m1} and 17% on σ_1 (absolute errors ~ 0.015 and ~ 0.011), whereas precision on the corresponding parameters of bin-7 are comparable to the other ‘no-cata’ bins. This is mainly due to the high- ℓ cut of bin-1 discard entirely the BAO signals, which can result in a 30 – 40% loss of precision according to Section 3.3. Concerning the catastrophic sub-sample in Bin-1(7), the mean redshift and width, z^c and σ^c , can be reconstructed with a fractional precision $\sim 6(19)\%$ and $\sim 50(40)\%$ respectively. The catastrophic fraction parameter f^c is constrained to a level of $\sim 40\%$ for both bins. f^c is considerably degenerated with the galaxy bias parameter b^c , due to the fact that they both impact the amplitude of the cross power spectrum.

The MCMC results indicate that it will be feasible in practical analysis to reconstruct the true redshift distribution of galaxies with angular power spectra. The precision on mean redshift and width of the tomographic bins can generally reach the level of $\sigma(z_{mi})/z_{mi} \sim 1\%$ and $\sigma(\sigma_i)/\sigma_i \sim 10\%$, respectively. To rewrite the uncertainty in z_m to be $\sigma(z_{mi})/(1 + z_{mi})$, which is then roughly equivalent to the bias in photo-z errors $e_z = (z_p - z_s)/(1 + z_s)$, it gives $\sim 0.005 - 0.008$ for different bins. Given the stringent requirement of a bias in e_z below 0.003 for a LSST-like survey to exploit its high statistical precision (e.g., LSST Science Collaboration et al. 2009), the spectroscopic sub-sample calibration and/or other cross correlation techniques should be combined to bridge the gap to the largest extent.

5 SUMMARY

The unprecedented statistical precision of future large weak lensing surveys require a rigorous control of systematic effects, such as those of photometric redshifts delivered by multi-band photometry. Hence precise calibration of the redshift distribution of the sampled galaxies, $n(z)$, is crucial in cosmological parameters estimation. The galaxy angular cross power spectra, determined by the true redshift overlap of two tomographic bins, are helpful for self-calibrating the photo-z bias and rms errors.

Considering an LSST-like fiducial survey, we investigate

⁶ <http://cosmologist.info/cosmomc/readme.html>

the contributions of various components of the galaxy angular power spectra in constraining the redshift distribution parameters. The principal role of cross spectra is confirmed, whereas the BAO feature is found to be relatively less important in this regard.

We further investigate, with MCMC analysis, the feasibility of this method in reconstructing the underlying redshift distribution when it is confronted with a practical data analysis. For the 10-bin fiducial survey, we find that the precision on mean redshift and width of the tomographic bins can generally reach the level of 1% and 10% ($\sigma(\theta_i)/\theta_i$), respectively. For bins with a catastrophic failure fraction, the mean redshift and width of the catastrophic sub-samples, z^c and σ^c , can be constrained to a fractional precision $\sim 5 - 20\%$ and $\sim 40 - 50\%$, respectively. Combined with other calibration methods such as the spectroscopic calibration, it is promising to approach the targeted precision on galaxy redshift distribution parameters.

For simplicity, we have ignored the intrinsic and the lensing-induced cross correlations between redshift bins. Although the simple double-Gaussian model suffices for this work, precision modeling of the underlying redshift distribution is of great importance. A principal component analysis directly on the simulated $P(z_s|z_p)$ map may be one of the next directions. With a focus on the performance of angular (cross) power spectra in reconstructing the galaxy redshift distributions, we fixed the cosmological parameters at their fiducial values. Nevertheless, the ultimate goal is to realize the self-calibration, namely, combining the weak lensing and galaxy clustering analysis on the very same galaxy sample to constrain both the cosmological and redshift distribution parameters, simultaneously.

ACKNOWLEDGEMENTS

We thank Qiao Wang and Teppei Okumura for useful discussions. This work was supported by the National Natural Science Foundation of China grant No. 11033005, the National Key Basic Research Science Foundation of China grant No. 2010CB833000, the Bairen program from the Chinese Academy of Sciences, and the Junior Researcher Foundation at the National Astronomical Observatories of China.

REFERENCES

- Abdalla F. B., Amara A., Capak P., Cypriano E. S., Lahav O., Rhodes J., 2008, *MNRAS*, **387**, 969
- Ahn S., Fessler J. A., 2003, EECS Department, The University of Michigan, pp 1–2
- Amara A., Réfrégier A., 2007, *MNRAS*, **381**, 1018
- Bacon D. J., Refregier A., Clowe D., Ellis R. S., 2001, *MNRAS*, **325**, 1065
- Bartelmann M., 2010, *Classical and Quantum Gravity*, **27**, 233001
- Benjamin J., Heymans C., Semboloni E., van Waerbeke L., et al. 2007, *MNRAS*, **381**, 702
- Benjamin J., van Waerbeke L., Ménard B., Kilbinger M., 2010, *MNRAS*, **408**, 1168
- Bridle S., et al., 2009, *Annals of Applied Statistics*, **3**, 6
- Cunha C. E., Lima M., Oyaizu H., Frieman J., Lin H., 2009, *MNRAS*, **396**, 2379
- Eisenstein D. J., Hu W., 1998, *ApJ*, **496**, 605
- Eisenstein D. J., Seo H.-J., White M., 2007, *ApJ*, **664**, 660
- Eriksen M., Gaztañaga E., 2015, *MNRAS*, **452**, 2149
- Fan Z.-H., 2007, *ApJ*, **669**, 10
- Fu L., et al., 2014, *MNRAS*, **441**, 2725
- Gilks W. R., Richardson S., Spiegelhalter D. J., 1996, Markov Chain Monte Carlo in Practice. Chapman and Hall, London
- Hearin A. P., Zentner A. R., Ma Z., Huterer D., 2010, *ApJ*, **720**, 1351
- Hearin A. P., Zentner A. R., Ma Z., 2012, *J. Cosmology Astropart. Phys.*, **4**, 34
- Heymans C., Brown M., Heavens A., Meisenheimer K., Taylor A., Wolf C., 2004, *MNRAS*, **347**, 895
- Heymans C., et al., 2006, *MNRAS*, **368**, 1323
- Hoekstra H., Yee H. K. C., Gladders M. D., 2002, *ApJ*, **577**, 595
- Hoekstra H., Mellier Y., van Waerbeke L., Semboloni E., et al. 2006, *ApJ*, **647**, 116
- Huterer D., Takada M., Bernstein G., Jain B., 2006, *MNRAS*, **366**, 101
- Jarvis M., Bernstein G. M., Fischer P., Smith D., Jain B., Tyson J. A., Wittman D., 2003, *AJ*, **125**, 1014
- Jee M. J., Tyson J. A., Schneider M. D., Wittman D., Schmidt S., Hilbert S., 2013, *ApJ*, **765**, 74
- Joachimi B., et al., 2015, *Space Sci. Rev.*,
- Kilbinger M., 2015, *Reports on Progress in Physics*, **78**, 086901
- Kilbinger M., Fu L., Heymans C., Simpson F., et al. 2013, *MNRAS*, **430**, 2200
- Köhlinger F., Viola M., Valkenburg W., Joachimi B., Hoekstra H., Kuijken K., 2015, preprint, ([arXiv:1509.04071](https://arxiv.org/abs/1509.04071))
- LSST Science Collaboration et al., 2009, preprint, ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- Lewis A., Bridle S., 2002, *Phys. Rev. D*, **66**, 103511
- Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, *MNRAS*, **390**, 118
- Liu X., et al., 2015, *MNRAS*, **450**, 2888
- Ma Z., Hu W., Huterer D., 2006, *ApJ*, **636**, 21
- Massey R., Rhodes J., Leauthaud A., Capak P., et al. 2007, *ApJS*, **172**, 239
- Matthews D. J., Newman J. A., 2010, *ApJ*, **721**, 456
- McQuinn M., White M., 2013, *MNRAS*, **433**, 2857
- Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint, ([arXiv:1303.4722](https://arxiv.org/abs/1303.4722))
- Miller L., Kitching T. D., Heymans C., Heavens A. F., van Waerbeke L., 2007, *MNRAS*, **382**, 315
- Rahman M., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., 2015, *MNRAS*, **447**, 3500
- Refregier A., 2003, *ARA&A*, **41**, 645
- Rhodes J., Refregier A., Collins N. R., Gardner J. P., Groth E. J., Hill R. S., 2004, *ApJ*, **605**, 29
- Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, *MNRAS*, **431**, 3307
- Schneider M., Knox L., Zhan H., Connolly A., 2006, *ApJ*, **651**, 14
- Semboloni E., Mellier Y., van Waerbeke L., Hoekstra H., et al. 2006, *A&A*, **452**, 51
- Sun L., Fan Z.-H., Tao C., Kneib J.-P., Jouvel S., Tilquin A., 2009, *ApJ*, **699**, 958
- Sun L., Wang Q., Zhan H., 2013, *ApJ*, **777**, 75
- Tegmark M., 1997, *Physical Review Letters*, **79**, 3806
- Yu Y., Zhang P., Lin W., Cui W., 2015, *ApJ*, **803**, 46
- Zhan H., 2006, *J. Cosmology Astropart. Phys.*, **8**, 8
- Zhan H., Knox L., 2006, *ApJ*, **644**, 663
- Zhan H., Knox L., Tyson J. A., Margoniner V., 2006, *ApJ*, **640**, 8
- Zhang P., Pen U.-L., Bernstein G., 2010, *MNRAS*, **405**, 359

This paper has been typeset from a \LaTeX file prepared by the author.